Draft of Solutions of Exam 2021
Multivariate Econometrics
VU Econometrics and Operations Research

This page is intentionally left blank.

Instructions

(i) All questions should be answered to get full points.

(ii) Each question is worth 25 points.

(iii) Read the instructions in the questions carefully.

(iv) Answer the questions as detailed as possible. Use mathematical expressions when necessary. You can use words when you cannot provide a formal mathematical answer to the questions.

(v) If a question is not clear to you, make your own assumptions to clarify the meaning of the question and then answer the question based on your assumptions.

(vi) See the back of this page for some standard results that you may make use of while answering the questions.

Some standard results

Suppose that the scalar process $\{z_t\}$ follows the following data generating process:

$$z_t = z_{t-1} + u_t,$$

where $z_0 = 0$ and u_t has the following properties:

(a) $u_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ where $\sum_{j=0}^{\infty} j \cdot |\psi_j| < \infty$ and $\{\epsilon_t\}$ is an $i.i.d$ sequence with mean zero and variance σ_ϵ^2, and finite fourth moment;

(b) σ^2 denotes the long run variance of $\{u_t\}$ and σ_u^2 denotes the contemporaneous variance of $\{u_t\}$.

Note that under these assumptions z_t can be written as a partial sum as

$$z_t = \sum_{s=1}^{t} u_s.$$

Let $W(r)$ be a standard Brownian motion process associated with u_t. Then the following results hold:

(1) $T^{-1/2} \sum_{t=1}^{T} u_t \xrightarrow{d} \sigma W(1)$;

(2) $T^{-1} \sum_{t=1}^{T} u_t^2 \xrightarrow{p} \sigma_u^2$;

(3) $T^{-1} \sum_{t=1}^{T} z_{t-1} u_t \xrightarrow{d} \frac{1}{2}\sigma^2 \left[W(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right]$;

(4) $T^{-3/2} \sum_{t=1}^{T} t u_{t-j} \xrightarrow{d} \sigma \left\{ W(1) - \int_0^1 W(r) dr \right\}$ for $j = 0, 1, \ldots$;

(5) $T^{-3/2} \sum_{t=1}^{T} z_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r) dr$;

(6) $T^{-2} \sum_{t=1}^{T} z_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 W(r)^2 dr$;

(7) $T^{-5/2} \sum_{t=1}^{T} t z_{t-1} \xrightarrow{d} \sigma \int_0^1 r W(r) dr$;

(8) $T^{-3} \sum_{t=1}^{T} t z_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 r W(r)^2 dr$;

(9) $T^{-(v+1)} \sum_{t=1}^{T} t^v \rightarrow 1/(v+1)$ for $v = 0, 1, \ldots$;

(10) Suppose that the DGP of another time series process y_t follows the model

$$y_t = y_{t-1} + e_t,$$

where $y_t = 0$ and e_t satisfy the same assumptions as **(a)** and has a long run variance σ_e^2 and $W_e(r)$ is a standard Brownian motion processs associated with e_t, then

$$T^{-2} \sum_{t=1}^{T} z_t y_t \xrightarrow{d} \sigma \sigma_e \int_0^1 W(r) W_e(r) dr.$$

Question 1: Conceptual Questions (25 points out of 100 points)

Below you will find **3** statements. All these are related to the concepts/techqniques that have been discussed during the lectures. Some of these statements are correct, some are wrong, some need further clarification. You need to provide a brief, to the point answer that would contain **(i) short explanations/definitions of the concepts mentioned in the statement, (ii) your judgement about the statement about whether it is correct/wrong/unclear/incomplete, and an explanation of your judgement (iii) a correction of the statement.** The concepts that you need to explain and define are written in *italics*. A formal answer using mathematics is possible, sometimes very useful but not always necessary.

(a) (5 points) Consider a *stable VAR(1) process*. The j^{th} order *autocovariance* of this process is equal to its 1^{th} order autocovariance, when $j \neq 1$.

(b) (10 points) Consider the time series x_t with the DGP

$$x_t = \lambda x_{t-1} + \varepsilon_t,$$

where $\lambda = 1$ and $\varepsilon_t \sim i.i.d(0, \sigma_\varepsilon^2)$. This implies that x_t has a *unit root*. IN this case, the OLS estimator of λ is $\sqrt{T}$-consistent and it is asymptotically normally distributed.

(c) (10 points) For panel data variables $y_{i,t}$ and $\mathbf{x}_{i,t}$ that are observed for N cross section units, over T time periods, a *homogeneous panel data model* can be written as

$$y_{i,t} = \boldsymbol{\beta}'\mathbf{x}_{i,t} + u_{i,t}.$$

When $Cov(u_{i,t}u_{j,t}) \neq 0$ for $i = j$, we say that i and j are cross-sectionally dependent. If we estimate $\boldsymbol{\beta}$ by the *pooled OLS estimator*, the estimator will be inconsistent. If we estimate $\boldsymbol{\beta}$ by using the *mean group estimator* the estimator will be consistent.

Answer:

(a) (5 points) Let $\mathbf{x}_t$ be an $m \times 1$ time series process which follows the VAR(1) model given below

$$\mathbf{x}_t = \boldsymbol{\lambda}'\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t.$$

In order to obtain the j^{th} order autocovariance of this procees, we need to rewrite it as

$$\mathbf{x}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Lambda}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Lambda}^2\boldsymbol{\varepsilon}_{t-2} + \cdots + \boldsymbol{\Lambda}^{j-1}\boldsymbol{\varepsilon}_{t-j+1} + \boldsymbol{\Lambda}^j\mathbf{x}_{t-j},$$

Then we can calculate the j^{th} order autocovariance as

$$\begin{aligned}
E(\mathbf{x}_t\mathbf{x}'_{t-j}) &= E[(\boldsymbol{\varepsilon}_t + \boldsymbol{\Lambda}\boldsymbol{\varepsilon}_{t-1} + \cdots + \boldsymbol{\Lambda}^{j-1}\boldsymbol{\varepsilon}_{t-j+1} + \boldsymbol{\Lambda}^j\mathbf{x}_{t-j})\mathbf{x}'_{t-j}] \\
&= E(\boldsymbol{\varepsilon}_t\mathbf{x}'_{t-j}) + \boldsymbol{\Lambda}E(\boldsymbol{\varepsilon}_{t-1}\mathbf{x}'_{t-j}) + \cdots \\
&\quad +\boldsymbol{\Lambda}^{j-1}E(\boldsymbol{\varepsilon}_{t-j+1}\mathbf{x}'_{t-j}) + \boldsymbol{\Lambda}^jE(\mathbf{x}_{t-j}\mathbf{x}'_{t-j}) \\
&\quad = \boldsymbol{\Lambda}^j\boldsymbol{\Omega}_x,
\end{aligned}$$

where $\boldsymbol{\Omega}_x$ is the variance covariance matrix of $\mathbf{x}$. $Cov(\mathbf{x}_t, \mathbf{x}'_{t-j})$ depends on j: the bigger the distance between $\mathbf{x}_t$, and $\mathbf{x}_{t-j}$ the smaller the covariance is. The statement is false, because as we see from the above result, the autocovariance matrix depends on j.

(b) (10 points) A unit root process is as given in the question

$$x_t = x_{t-1} + \varepsilon_t.$$

This process is sometimes called the random walk process. Estimating models that have unit root leads to different results than the standard stationary case. The statement is false, because $\sqrt{T}$ consistency rate is the consistency rate of the OLS estimator in the stationary case. If we estimate λ when $\lambda = 1$ by simple OLS estimator, the estimator will be $T-$consistent. This is also called super consistency. Because the estimator will converge to its true value faster. The asymptotic distribution of this estimator is not normal. It is a functional of Brownian motions. As there is no serial correlation in the errors, the limiting distribution will be nuisance parameter free. The critical values are tabulated by Dickey-Fuller.

(c) (10 points) Consider the following model

$$y_{i,t} = \boldsymbol{\beta}_i' \mathbf{x}_{i,t} + u_{i,t},$$

This model is a linear heterogeneous panel data model. A heterogeneous panel data model is a model in which the parameters vary across individuals. A homogeneous panel data model is a model in which all parameters are the same for all cross-section units: $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for all $i = 1, \ldots, N$. We can estimate the $\boldsymbol{\beta}$, which is defined in Assumption 1, by

$$\widehat{\boldsymbol{\beta}}_P = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{y}_i,$$

We can write the same estimator using the summation notation as

$$\widehat{\boldsymbol{\beta}}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\beta}}_i = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{y}_i$$

When there is cross-sectional dependence, the effect of it on the estimators depend on the extend of the cross-sectional dependence. If this cross-sectional dependence leads to endogeneity then both pooled and mean group estimators will be inconsistent. A situation where cross-sectional dependence leads to endogeneity is when there are unobserved global factors affecting $y_{i,t}$ and $\mathbf{x}_{i,t}$. In this case the error term $u_{i,t}$ will be correlated with the regressors. Then the estimators will be inconsistent. So the statement is neither false not true. The answer depends on the extend of cross-sectional dependence. Also the definition of cross-sectional dependence given in the question is not correct as it is written that $Cov(u_{i,t} u_{j,t}) \neq 0$ for $i = j$. It should be $Cov(u_{i,t} u_{j,t}) \neq 0$ for $i \neq j$

Question 2: Modeling and stationarity (25 points out of 100 points)

(a) (15 points) Consider three processes y_t, x_t and z_t with the DGPs

$$\Delta y_t = \alpha_1(y_{t-1} - \beta_1 x_{t-1} - \beta_2 z_{t-1}) + \gamma_{11}\Delta y_{t-1} + \gamma_{13}\Delta z_{t-1} + u_{y,t},$$
$$\Delta x_t = \alpha_2(y_{t-1} - \beta_1 x_{t-1} - \beta_2 z_{t-1}) + \gamma_{21}\Delta y_{t-1} + \gamma_{22}\Delta x_{t-1} + u_{x,t},$$
$$\Delta z_t = \alpha_3(y_{t-1} - \beta_1 x_{t-1} - \beta_2 z_{t-1}) + \gamma_{32}\Delta x_{t-1} + \gamma_{33}\Delta z_{t-1} + u_{z,t},$$

where

$$\begin{pmatrix} u_{y,t} \\ u_{x,t} \\ u_{z,t} \end{pmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}\right].$$

Now we define

$$\mathbf{w}_t = \begin{pmatrix} y_t \\ x_t \\ z_t \end{pmatrix}.$$

(i) Write the vector error correction model (VECM) for $\Delta\mathbf{w}_t$.

(ii) Obtain the vector autoregression representation (VAR) for $\mathbf{w}_t$.

(iii) Assume that $\alpha_2 = \alpha_3 = \gamma_{21} = \gamma_{32} = 0$ and $|\gamma_{22}| < 1$, $|\gamma_{33}| < 1$. Establish and discuss the order of integration d, i.e. I(d) of y_t, x_t and z_t and the cointegration relations between them (as well as the number and the form of the cointegrating vectors, if any).

(iv) Derive the conditional error correction model (CECM) for y_t given x_t and its past (you can ignore the model for z_t).

(b) (10 points) For panel data variables $y_{i,t}$ and $\mathbf{x}_{i,t}$ that are observed for N cross section units, over T time periods, a homogeneous panel data model can be written as

$$y_{i,t} = \boldsymbol{\beta}'\mathbf{x}_{i,t} + u_{i,t},$$

where

$$u_{i,t} = \boldsymbol{\lambda}_i'\mathbf{f}_t + \varepsilon_{i,t},$$

where $\mathbf{f}_t$ is a $k \times 1$ vector of unobserved global shocks, $\boldsymbol{\lambda}_i$ is the $k \times 1$ vector of factor loadings and $\varepsilon_{i,t}$ is the idiosyncratic error. Furthermore, we assume that the $m \times 1$ vector of $\mathbf{x}_{i,t}$ has the following DGP:

$$\mathbf{x}_{i,t} = \boldsymbol{\Lambda}_i\mathbf{f}_t + \boldsymbol{\nu}_{i,t},$$

where $\boldsymbol{\Lambda}_i$ is the $m \times k$ matrix that contains the factor loadings for $\mathbf{x}_{i,t}$. Now, define the $m+1$ vector $\mathbf{z}_{i,t}$ as

$$\mathbf{z}_{i,t} = \begin{pmatrix} y_{i,t} \\ \mathbf{x}_{i,t} \end{pmatrix}.$$

For this $\mathbf{z}_{i,t}$ we can obtain a factor model of the form

$$\mathbf{z}_{i,t} = \mathbf{C}_i\mathbf{f}_t + \mathbf{e}_{i,t}.$$

Obtain this factor model by using the models for $y_{i,t}$ and $\mathbf{x}_{i,t}$ and write $\mathbf{C}_i$ and $\mathbf{e}_{i,t}$ in terms the model components of the models for $y_{i,t}$ and $\mathbf{x}_{i,t}$.

(a) (15 points) Here (i) is 3 points, (ii) is 10 points, (iii) is 2 points.

(i) (2 points) We need to obtain the VECM representation. Single equation representation is already given. We can define the following matrices and vectors to write the models in the VECM form.

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} 1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad \boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & 0 & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & 0 \\ 0 & \gamma_{32} & \gamma_{33} \end{pmatrix},$$

$$\mathbf{w}_t = \begin{pmatrix} y_t \\ x_t \\ z_t \end{pmatrix}, \quad \mathbf{u}_t = \begin{pmatrix} u_{y,t} \\ u_{x,t} \\ u_{z,t} \end{pmatrix},$$

Now the VECM can be written as

$$\Delta \mathbf{w}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{w}_{t-1} + \boldsymbol{\Gamma}\Delta\mathbf{w}_{t-1} + \mathbf{u}_t.$$

(ii) (3 points) Obtaining the VAR model requires rearrangement of the VECM model. We can write

$$\mathbf{w}_t - \mathbf{w}_{t-1} = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{w}_{t-1} + \boldsymbol{\Gamma}\mathbf{w}_{t-1} - \boldsymbol{\Gamma}\mathbf{w}_{t-2} + \mathbf{u}_t$$
$$\mathbf{w}_t = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{w}_{t-1} + \mathbf{w}_{t-1} + \boldsymbol{\Gamma}\mathbf{w}_{t-1} - \boldsymbol{\Gamma}\mathbf{w}_{t-2} + \mathbf{u}_t$$

Grouping the terms gives

$$\mathbf{w}_t = (\boldsymbol{\alpha}\boldsymbol{\beta}' + \mathbf{I}_3 + \boldsymbol{\Gamma})\mathbf{w}_{t-1} - \boldsymbol{\Gamma}\mathbf{w}_{t-2} + \mathbf{u}_t$$

So the VAR model is

$$\mathbf{w}_t = \mathbf{A}\mathbf{w}_{t-1} + \mathbf{B}\mathbf{w}_{t-2} + \mathbf{u}_t,$$

where $\mathbf{A} = \boldsymbol{\alpha}\boldsymbol{\beta}' + \mathbf{I}_3 + \boldsymbol{\Gamma}$ and $\mathbf{B} = -\boldsymbol{\Gamma}$.

(iii) (5 points) We can rewrite the model based on the restrictions given in the question. This will give

$$\Delta y_t = \alpha_1(y_{t-1} - \beta_1 x_{t-1} - \beta_2 z_{t-1}) + \gamma_{11}\Delta y_{t-1} + \gamma_{13}\Delta z_{t-1} + u_{y,t},$$
$$\Delta x_t = \gamma_{22}\Delta x_{t-1} + u_{x,t},$$
$$\Delta z_t = \gamma_{33}\Delta z_{t-1} + u_{z,t},$$

where $|\gamma_{22}| < 1$, $|\gamma_{33}| < 1$.

We can start with determining the order of integration of z_t. As $|\gamma_{33}| < 1$, we know that $\Delta z_t \sim I(0)$. This makes z_t $I(1)$. Same holds for x_t. Now we have established that x_t and z_t are $I(1)$. When we look at the model for Δy_t we see that x_t and z_t are in the model. There are two possibilities for y_t then, either y_t is $I(1)$ and x_t, z_t and y_t are cointegrated with the cointegrating vector $(1, \beta_1, \beta_2)$. Or y_t is $I(0)$ and only z_t and x_t are cointegrated with the cointegrating vector (β_1, β_2). For the VECM to be stable α_1 needs to be in between -2 and 0.

(iv) (5 points) Derivation of the error correction model is a standard exercise. Writing the conditional ECM model: We start with writing the models for Δy_t and Δx_t in a VECM form under given restrictions.

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} (y_{t-1} - \beta_1 x_{t-1} - \beta_2 z_{t-1}) + \begin{pmatrix} \gamma_{11} & 0 & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & 0 \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix}$$

In order to obtain a conditional model, we need to make sure that the conditional error term is independent of the error term of the model of conditioning variables. Let us denote the conditional error term by $u_{y.x,t}$. We need to have $E(u_{y.x,t}u_{x,t}) = 0$. How can we construct such an error term? We would like to construct $u_{y.x,t}$ such that $E(u_{y.x,t}u_{x,t}) = 0.$. We know that

$$E \begin{pmatrix} u_{y,t}^2 & u_{y,t}u_{x,t} \\ u_{x,t}u_{y,t} & u_{x,t}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

We can construct $u_{y,x,t} = u_{y,t} - \sigma_{21}\sigma_{22}^{-1}u_{x,t}$. This is derived in the exercise set Part II. Then we have

$$E(u_{y.x,t}u_{x,t}) = E(u_{y,t}u_{x,t}) - \sigma_{21}\sigma_{22}^{-1}E(u_{x,t}u_{y,t}) = 0$$

So in order to obtain this conditional error term we need to premultiply the vector $\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$ by $(1, -\sigma_{21}\sigma_{22}^{-1})$ (Note that $\sigma_{21} = \sigma_{12}$). We now consider the

VECM model again and premultiply both sides by $(1, -\sigma_{21}\sigma_{22}^{-1})$.

$$\begin{aligned}&\begin{pmatrix} 1 & -\sigma_{21}\sigma_{22}^{-1} \end{pmatrix}\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} \\ =&\begin{pmatrix} 1 & -\sigma_{21}\sigma_{22}^{-1} \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}(y_{t-1} - \beta_1 x_{t-1} - \beta_2 z_{t-1}) \\ &+\begin{pmatrix} 1 & -\sigma_{21}\sigma_{22}^{-1} \end{pmatrix}\begin{pmatrix} \gamma_{11} & 0 & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & 0 \end{pmatrix}\begin{pmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \\ \Delta z_{t-1} \end{pmatrix} \\ &+\begin{pmatrix} 1 & -\sigma_{21}\sigma_{22}^{-1} \end{pmatrix}\begin{pmatrix} u_{y,t} \\ u_{x,t} \end{pmatrix}\end{aligned}$$

This gives:

$$\begin{aligned}\Delta y_t &= (\alpha_1 - \sigma_{21}\sigma_{22}^{-1}\alpha_2)(y_{t-1} - \beta_1 x_{t-1} - \beta_2 z_{t-1}) \\ &+ \sigma_{21}\sigma_{22}^{-1}\Delta x_t + \gamma_{11}\Delta y_{t-1} - \sigma_{21}\sigma_{22}^{-1}\gamma_{22}\Delta x_{t-1} + \gamma_{13}\Delta z_{t-1} + u_{y.x,t}\end{aligned}$$

This is the conditional ECM of y_t given z_t and the past.

(b) (10 points) This question is rewriting a panel data model to write it in a factor model form. Consider the models for $y_{i,t}$ and $\mathbf{x}_{i,t}$.

$$\begin{aligned}y_{i,t} &= \boldsymbol{\beta}_i'\mathbf{x}_{i,t} + \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{i,t}, \\ \mathbf{x}_{i,t} &= \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{i,t}.\end{aligned}$$

Replace $\mathbf{x}_{i,t}$ with $\boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{i,t}$ in the model for $y_{i,t}$.

$$\begin{aligned}y_{i,t} &= \boldsymbol{\beta}_i'(\boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{i,t}) + \boldsymbol{\gamma}_i'\mathbf{f}_t + \varepsilon_{i,t}, \\ \mathbf{x}_{i,t} &= \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{i,t}.\end{aligned}$$

Rearrange it

$$\begin{aligned}y_{i,t} &= (\boldsymbol{\beta}_i'\boldsymbol{\Gamma}_i' + \boldsymbol{\gamma}_i')\mathbf{f}_t + \boldsymbol{\beta}_i'\mathbf{v}_{i,t} + \varepsilon_{i,t}, \\ \mathbf{x}_{i,t} &= \boldsymbol{\Gamma}_i'\mathbf{f}_t + \mathbf{v}_{i,t}.\end{aligned}$$

By using these resulting models from the previous slide we can write a model for $\mathbf{z}_{i,t} = (y_{i,t}, \mathbf{x}_{i,t}')'$:

$$\mathbf{z}_{i,t} = \mathbf{C}_i'\mathbf{f}_t + \mathbf{u}_{i,t},$$

where

$$\mathbf{u}_{i,t} = \begin{pmatrix} \varepsilon_{i,t} + \boldsymbol{\beta}'\mathbf{v}_{i,t} \\ \mathbf{v}_{i,t} \end{pmatrix},$$

and

$$\mathbf{C}_i = \begin{pmatrix} \boldsymbol{\gamma}_i + \boldsymbol{\Gamma}_i \boldsymbol{\beta}_i & \boldsymbol{\Gamma}_i \end{pmatrix},$$

here $\mathbf{C}_i$ is a $m \times (k+1)$ matrix.

Question 3: Asymptotic Derivations (25 points out of 100 points)

(a) (15 points) **Costs of ignoring constants and trends in the cointegrating regression:**

Suppose that we have the following data generating processes (DGP) for $\{y_t\}$ and $\{x_t\}$

$$y_t = \delta + \beta x_t + u_{1,t},$$
$$x_t = \mu + x_{t-1} + u_{2,t},$$

for $t = 1, \ldots, T$. We assume:

- $u_{1,t} = \psi_1(L)\epsilon_{1,t} = \sum_{j=0}^{\infty} \psi_{1,j}\epsilon_{1,t-j}$ where $\sum_{j=0}^{\infty} j \cdot |\psi_{1,j}| < \infty$ and $\{\epsilon_{1,t}\}$ is an $i.i.d$ sequence with mean zero and variance $\sigma_{\epsilon,1}^2$, and finite fourth moment, σ_1^2 denotes the long run variance of $\{u_{1,t}\}$ and $\sigma_{u,1}^2$ denotes the contemporaneous variance of $\{u_{1,t}\}$;
- $u_{2,t} = \psi_2(L)\epsilon_{2,t} = \sum_{j=0}^{\infty} \psi_{2,j}\epsilon_{2,t-j}$ where $\sum_{j=0}^{\infty} j \cdot |\psi_{2,j}| < \infty$ and $\{\epsilon_{2,t}\}$ is an $i.i.d$ sequence with mean zero and variance $\sigma_{\epsilon,2}^2$, and finite fourth moment, σ_2^2 denotes the long run variance of $\{u_{2,t}\}$ and $\sigma_{u,2}^2$ denotes the contemporaneous variance of $\{u_{2,t}\}$;
- $y_0 = x_0 = 0$.

A researcher wants to estimate β and for this purpose considers the following regression model

$$y_t = \beta x_t + error.$$

The OLS estimator is given by

$$\widehat{\beta} = \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \sum_{t=1}^{T} x_t y_t.$$

The error of the estimator can be obtained by plugging in the model for y_t into this estimator. This will result in

$$\widehat{\beta} - \beta = \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \sum_{t=1}^{T} x_t(\delta + u_{1,t}).$$

(i) Derive and describe the order of probability and the asymptotic distribution of $\widehat{\beta} - \beta$. Write down any necessary additional assumptions (if any).

(ii) Comment on your results. Is there a problem with this estimator?

(b) (10 points) **Costs of ignoring the time trend in the unit root AR(1) regression:**

Suppose that we have the following data generating process (DGP) for $\{y_t\}$

$$y_t = \delta + y_{t-1} + u_t$$

for $t = 1, \ldots, T$. We assume:

- $u_t = \psi(L)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ where $\sum_{j=0}^{\infty} j \cdot |\psi_j| < \infty$ and $\{\epsilon_t\}$ is an $i.i.d$ sequence with mean zero and variance σ_ϵ^2, and finite fourth moment. σ_y^2 denotes the long run variance of $\{u_t\}$ and σ_u^2 denotes the contemporaneous variance of $\{u_t\}$;
- $y_0 = 0$;

We consider the estimation of the regression model

$$y_t = \lambda y_{t-1} + error,$$

using a sample of T observations.

Consider the least squares estimator

$$\widehat{\lambda} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}.$$

The error of the estimator can be obtained by plugging in the model for y_t into this estimator. This will result in

$$\widehat{\lambda} - 1 = \left(\sum_{t=1}^{T} y_{t-1}^2 \right)^{-1} \sum_{t=1}^{T} y_{t-1}(\delta + u_{y,t}).$$

(i) Derive and discuss the orders of probability and limiting distributions of the numerator and the denominator of $(\widehat{\lambda} - 1)$. Write down any necessary additional assumptions (if any).

(ii) Comment on your results. Is there a problem with this estimator?

Answer:

(a) (15 points)

(i) (10 points) We need to derive and describe the order of probability and the asymptotic distribution of

$$\widehat{\beta} - \beta = \left(\sum_{t=1}^{T} x_t^2\right)^{-1} \sum_{t=1}^{T} x_t(\delta + u_{1,t}).$$

Let's start with the denominator. x_t is a random walk with a drift. So it can be rewritten as

$$x_t = \mu t + S_t,$$

where $S_t = \sum_{s=1}^{t} u_{2,t}$. We can apply all the standard results to S_t. Rewriting the denominator gives

$$\sum_{t=1}^{T} x_t^2 = \sum_{t=1}^{T} (\mu t + S_t)^2 = \mu^2 \sum_{t=1}^{T} t^2 + \mu \sum_{t=1}^{T} t S_t + \sum_{t=1}^{T} S_t^2.$$

According to standard result 6, the third term is $O_p(T^2)$. According to standard result 7, the second term is $O_p(T^{5/2})$, according to standard result 9, the first term is $O_p(T^3)$. Then the first term is the dominant term. This gives

$$\frac{1}{T^3} \sum_{t=1}^{T} x_t^2 = \mu^2 \frac{1}{T^3} \sum_{t=1}^{T} t^2 + o_p(1)$$

and

$$\frac{1}{T^3} \sum_{t=1}^{T} x_t^2 \to_p \frac{\mu^2}{3}.$$

Now we will obtain the limit of the numerator. We have

$$\sum_{t=1}^{T} x_t(\delta + u_{1,t}) = \sum_{t=1}^{T} (\mu t + S_t)(\delta + u_{1,t}) = \mu\delta \sum_{t=1}^{T} t + \mu \sum_{t=1}^{T} t u_{1,t} + \delta \sum_{t=1}^{T} S_t + \sum_{t=1}^{T} S_t u_{1,t}$$

According to standard result 3, the fourth term is $O_p(T)$. According to standard result 5, the third term is $O_p(T^{3/2})$, according to standard result 4, the first term is $O_p(T^{3/2})$. Finally, according to standard result 9, the first term is $O_p(T^2)$. Then the first term is the dominant term. This gives

$$\frac{1}{T^2} \sum_{t=1}^{T} x_t(\delta + u_{1,t}) = \frac{1}{T^2} \mu\delta \sum_{t=1}^{T} t + o_p(1)$$

and

$$\frac{1}{T^2}\sum_{t=1}^{T} x_t(\delta + u_{1,t}) \to_p \frac{\mu\delta}{2}.$$

The limiting behaviour of $\widehat{\beta} - \beta$ is given by

$$T(\widehat{\beta} - \beta) \to_p \frac{3\delta}{2\mu}.$$

(ii) (5 points) Here we are ignoring the constant from the cointegrating regression. This question is about understanding the effects of this. In part i, we found that $\widehat{\beta} - \beta) \to_p 0$. So, $\widehat{\beta}$ is a consistent estimator of β. However there is a problem with the limiting distribution. When we rescale the $\widehat{\beta} - \beta$ with T, the limit is a constant and is a function of μ and δ. We need a random variable in the limit to make inference and also the random limit should not be function of nuisance parameters such as μ and δ. This analysis makes us conclude that either ignoring the constant or x_t being a random walk with a drift or both causes problem in the inference of the cointegrating vector.

(b) (10 points)

(i) (7 points)We need to derive and discuss the orders of probability and limiting distributions of the numerator and the denominator of

$$\widehat{\lambda} - 1 = \left(\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} \sum_{t=1}^{T} y_{t-1}(\delta + u_{y,t}).$$

We start with the denominator. y_t is a random walk with a drift so it can be written as

$$y_t = \delta t + S_t,$$

where $S_t = \sum_{s=1}^{t} u_t$ is the partial sum that we can apply the standard results. We can now rewrite the denominator as

$$\sum_{t=1}^{T} y_{t-1}^2 = \sum_{t=1}^{T}(\delta(t-1) + S_{t-1})^2 = \delta^2\sum_{t=1}^{T}(t-1)^2 + \delta\sum_{t=1}^{T}(t-1)S_{t-1} + \sum_{t=1}^{T} S_{t-1}^2.$$

According to standard result 6, the third term is $O_p(T^2)$. According to standard result 7, the second term is $O_p(T^{5/2})$, according to standard result 9, the first term is $O_p(T^3)$. Then the first term is the dominant term. This gives

$$\frac{1}{T^3}\sum_{t=1}^{T} y_{t-1}^2 = \delta^2 \frac{1}{T^3}\sum_{t=1}^{T}(t-1)^2 + o_p(1)$$

and

$$\frac{1}{T^3}\sum_{t=1}^{T} y_{t-1}^2 \to_p \frac{\delta^2}{3}.$$

Now we will obtain the limit of the numerator. We have

$$\begin{aligned}\sum_{t=1}^{T} y_{t-1}(\delta + u_{y,t}) &= \sum_{t=1}^{T}[\delta(t-1) + S_{t-1}](\delta + u_{y,t}) \\ &= \delta^2\sum_{t=1}^{T}(t-1) + \delta\sum_{t=1}^{T}(t-1)u_{y,t} + \delta\sum_{t=1}^{T} S_{t-1} + \sum_{t=1}^{T} S_{t-1}u_{y,t}\end{aligned}$$

According to standard result 3, the fourth term is $O_p(T)$. According to standard result 5, the third term is $O_p(T^{3/2})$, according to standard result 4, the first term is $O_p(T^{3/2})$. Finally, according to standard result 9, the first term is $O_p(T^2)$. Then the first term is the dominant term. This gives

$$\sum_{t=1}^{T} y_{t-1}(\delta + u_{y,t}) = \frac{1}{T^2}\delta^2\sum_{t=1}^{T}(t-1) + o_p(1)$$

and

$$\frac{1}{T^2}\sum_{t=1}^{T} u_t(\delta + u_{y,t}) \to_p \frac{\delta^2}{2}.$$

(ii) (3 points) The findings of part i, tells us that the limiting behaviour of the error of the estimator is as follows:

$$T(\widehat{\lambda} - 1) = \left(\frac{1}{T^3}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} \frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}(\delta + u_{y,t}) \to_p \frac{2}{3}.$$

This tells us that $\widehat{\lambda}$ converges to 1. However the scaled version is converging to a constant that is 2/3. This is not good for inference purposes. When we ignore the drift, it means that the OLS estimator can not be used for testing hypothesis.

Question 4: Empirical Application (25 points out of 100 points)

(a) (15 points) Bruce Wayne is an econometrics student at Gotham City University. He wants to analyze the price elasticity of Batmobiles. He collects annual data on the following variables for the period 1921 - 2021:

- log of prices of Batmobiles ($price_t$),
- log of annual sales quantity of Batmobiles ($quant_t$)
- log of interest rates (int_t)
- log of annual inflation in Gotham (inf_t)

He knows from previous empirical research that all these variables contain unit roots. He wants to estimate long run relations between the variables. He starts his analysis by considering the trace test of Johansen. He produces the following output.

Null hypothesis	t-stat	5% critical values
$r = 0$	80.21	69.81
$r = 1$	64.44	47.85
$r = 2$	50.46	29.79
$r = 3$	10.27	15.49

Here r is the cointegration rank. Then, he considers a single equation regression to estimate the long run relation between these variables, he estimates by OLS the following regression

$$price_t = \beta_1 quant_t + \beta_2 int_t + \beta_3 inf_t + error.$$

(i) What conclusions can you draw from the results of the trace test.

(ii) Is there are problem with Bruce's approach of estimating a single static least squares regression? Explain in detail.

(iii) You know that he wants to estimate long run relations between the variables. What are your suggestions to Bruce? Explain in detail.

(b) (10 points) Harvey Dent is another econometrics student at Gotham City University. He is analyzing a panel data set of crime rates of 200 towns over 5 years. Let $c_{i,t}$ denote the number of crimes committed in town i. Previous research shows that crime rates data is dynamic in nature and it is highly suspected that an individual specific effect is present in the error term of the model that represents the heterogeneity across towns. Harvey realizes these and assumes the following model, where he denotes the variable by $c_{i,t}$:

$$c_{i,t} = \lambda c_{i,t-1} + u_{i,t},$$

where $u_{i,t} = \mu_i + \varepsilon_{i,t}$. He starts looking for an advice on how to analyze this panel data set.

(i) One of the ways to estimate this model is to use a fixed effects (within transformation) approach. However, it is known to be problematic. Explain why the fixed effects approach is problematic.

(ii) Propose and discuss briefly an alternative estimator that is more proper/optimal than the fixed effects estimator.

Answer:

(a) (15 points)

- **(i)** (5 points) Here you need to discuss what trace test is, what is the procedure we need to follow to interpret a trace test. Also you need to explain the purpose of doing a trace test. Finally you need to write your conclusions from the results of the trace test.
- **(ii)** (5 points) Trace test gives evidence for multiple cointegrating relations. But Bruce continues with a single equation estimation. This is problematic. You need to explain why this is a problem.
- **(iii)** (5 points) Here you need to suggest a VECM approach of Johansen and explain the ML estimation of VECM proposed by Johansen.

(b) (10 points)

- **(i)** (5 points) The problem with estimating a dynamic panel data model with fixed effects is problematic because of the Nickell bias. Here, you need to explain the problem of Nickell bias with all the details.
- **(ii)** (5 points) The alternatives are Anderson and Hsiao, Arrelano and Bond. Here you need to explain these methods with all the details. Discuss their advantages and disadvantages.